β-Thalassemia Due to Two Novel Nucleotide Substitutions in Consensus Acceptor Splice Sequences of the β-Globin Gene

By Corinne Wong, Stylianos E. Antonarakis, Sabra C. Goff, Stuart H. Orkin, Bernard G. Forget, David G. Nathan, Patricia J.V. Giardina, and Haig H. Kazazian, Jr

We have identified two novel RNA-splicing mutations affecting a critical nucleotide (nt) in the acceptor consensus sequences at both the IVS-1/exon 2 and IVS-2/exon 3 junctions of the human β-globin gene. Both mutations are single nt substitutions, T to G and C to A, at position −3 adjacent to the invariant AG dinucleotide. For the IVS-2/exon 3 mutation abnormal splicing into the cryptic splice site at IVS-2 nt 579 is documented. Identification of these two mutations provides further support for the importance of the location of specific nucleotides within the consensus sequences in splice site selection and RNA processing.

COMPILATIONS of splice junction sequences from viral and eukaryotic genes[1,2] have revealed well-defined consensus sequences at all exon/intron boundaries. These observations include the invariant dinucleotide pairs, GT and AG, which demark the beginnings and ends of introns, respectively.[3] By alignment of sequences according to the GT/AG rule and tabulation of nucleotide frequencies by position, donor sequences conform to the consensus, $^{C}_{A}AG/GT^{A}_{G}AGT$, while the acceptor consensus sequence is preceded by a long pyrimidine stretch, $(^{T}_{C})_{11}N^{T}_{C}AG/G$.[1,2] The nonrandomness and extreme evolutionary conservation of these consensus sequences suggests a precise RNA splicing mechanism and one that is functionally dependent on conformation to these rules.[1] Beyond simple sequence comparisons, the evidence implicating these consensus sequences in proper RNA splicing has accumulated in vivo with naturally occurring pathologic defects producing β-thalassemia,[4] hemophilia B,[5] and phenylketonuria.[6] In addition, transcriptional and expression analyses of these genes after introduction into HeLa cells has provided further support for the importance of these sequences.[4]

Nearly 50% of the naturally occurring point mutations producing β-thalassemia have been shown to affect RNA splicing. No mutations, however, have been described in the consensus sequences outside the invariant dinucleotides AG or GT at two of the four splice junctions of the β-globin gene, namely, the IVS-1 acceptor site and IVS-2 donor site. A single nucleotide substitution at position −6 before the AG sequence at the acceptor site of IVS-2 has been reported, but functional studies on this mutation have not yet been carried out.[7] We now report two mutations involving alterations of both acceptor consensus sequences. The first is a single C to A substitution in the acceptor consensus sequence of the second intron at position −3, observed in a β-globin gene in both an Iranian and an Egyptian patient with β-thalassemia. The second mutation is a T to G change also at position −3 of the acceptor site but in the first intron of the β-globin gene. This latter mutation was identified in a Saudi Arabian individual by direct genomic sequencing. Analysis of the IVS-2 acceptor site mutation in transient expression studies suggests that the nucleotide −3 of the acceptor consensus sequence is quantitatively critical for correct splicing and that there exists a nucleotide sequence dependent component in RNA splicing.

From the Genetics Unit, Department of Pediatrics, The Johns Hopkins University School of Medicine, Baltimore; Howard Hughes Medical Institute, Division of Hematology-Oncology, Children's Hospital and the Dana-Farber Cancer Institute, Department of Pediatrics, Harvard Medical School, Boston; Hematology Section, Department of Internal Medicine, Yale University School of Medicine, New Haven; the Department of Pediatrics, Children's Hospital, Harvard Medical School, Boston; and the Division of Pediatric Hematology-Oncology, New York Hospital and Cornell Medical Center.

Submitted August 9, 1988; accepted October 27, 1988.

Supported by National Institutes of Health (NIH) grants to H.H.K., S.E.A., and S.H.O., and MOD grants to H.H.K. and S.E.A. C.W. was supported by an NIH Predoctoral Training Grant in Human Genetics.

Address reprint requests to Haig Kazazian, MD, Department of Pediatrics, CMSC 10-110, Johns Hopkins Hospital, 600 N Wolfe St, Baltimore, MD 21205.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" *in accordance with 18 U.S.C. section 1734 solely to indicate this fact.*

0006-4971/89/7304-0001$3.00

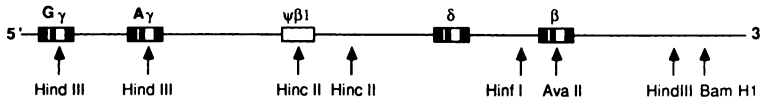
MATERIALS AND METHODS

Subjects. The Iranian patient had severe transfusion-dependent β-thalassemia and was a genetic compound with two different mutant alleles of the β-globin gene. The β-thalassemia trait parent who carried the acceptor site mutation had a mean corpuscular volume (MCV) of 76 fL and a hemoglobin A_2 of 5.5%. The Egyptian patient presented clinically with β-thalassemia trait. He and his wife, who carried the same mutation, had MCVs of 69 fL and 73 fL and hemoglobin A_2 values of 3.9% and 4.1%. The last patient of Saudi Arabian background was a genetic compound with transfusion-dependent B^+-thalassemia major. Sequence analysis of parental DNA revealed that this patient inherited an IVS-1 position 110 mutation from his mother and the acceptor site mutation from his father. His father has an MCV of 70 fL and hemoglobin A_2 of 4.6%.

DNA polymorphism analysis. Eight polymorphic restriction endonuclease sites described previously,[8,9] within and surrounding the β-globin gene, were used to define the frameworks and haplotypes as shown in Fig 1.

Genomic cloning. DNA isolated from peripheral blood leukocytes from the Iranian patient was digested with *Hind* III and the 7 to 9 kb size-selected fragments were cloned into the vector Charon 28.[9] The partial library was screened with a *Bam* HI-*Eco* RI fragment encompassing IVS-2 of the β-globin gene.

M13 DNA sequencing. *BGl* II/*Pst* I subclones, 3.7 kb, encompassing the entire β-gene from the Iranian patient were constructed in M13 for dideoxynucleotide sequencing.[10] Eight synthetic 19mer oligonucleotides complementary to the regions around (*a*) nucleotide −270 5′ to the β-globin gene, (*b*) nucleotide −28 in the TATA box, (*c*) IVS-1 nt 1, (*d*) codon 71 in exon 2, (*e*) IVS-2 nt 160, (*f*)



a) Haplotype VII Med Fa Ind	–		–		–	–		+	–	–	+
Framework 3											
b) Haplotype XI Med	+		+		–	–		+	+	+	–
Framework 2											
c) Haplotype VII Med	–		–		–	–		+	–	–	+
Framework 3a											

Fig 1. Haplotypes were constructed by family studies using eight polymorphic restriction endonuclease sites within the β-globin gene cluster as described previously.[8] a, Iranian allele; b, Egyptian allele: c, Saudi Arabian allele.

IVS-2 nt 400, (*g*) IVS-2 nt 666, and (*h*) codon 110 in exon 3 were used as primers for sequencing the entire β-globin gene.

PCR amplification and direct genomic sequencing. The β-globin genes from the Egyptian and Saudi Arabian patients were enzymatically amplified by the polymerase chain reaction (PCR).[11] Two sets of PCR primers encompassing a 770 base pair (bp) region from position −160 of the β-gene to 100 nucleotides with IVS-2, and a 564 bp region including exon 3 and the polyadenylation sequence were used. Genomic DNA (0.5 μg) was subjected to 30 cycles of PCR using Taq polymerase with a 30 second denaturing period at 94°C, a 30 second annealing period at 55°C, and a 3 minute extension period at 72°C. Amplified material was directly sequenced as reported[12] using third sequencing primers as described for M13 sequencing. Briefly, end-labeled oligonucleotide primers were annealed to denatured DNA and synthesis was carried out with Moloney murine leukemia virus (MMLV)-reverse transcriptase and cold nucleotides.

Transient expression assay in vitro. The β-gene from the Iranian patient was subcloned as a 3.7 kb *Bgl* II-*Pst* I fragment into the transient expression vector πSVplac.[4] The β-gene construct was transferred into HeLa cells and total cellular RNA was extracted 48 hours posttransfection for S1 mapping and primer extension experiments.[4]

RESULTS

β-globin gene haplotype and framework analysis. Haplotype analysis of the Iranian alleles revealed a haplotype VII Mediterranean or Fa Indian as shown in Fig 1A (for haplotype nomenclature, see reference 13). Previous examination of this patient's other β-globin gene cluster revealed a haplotype I Mediterranean or haplotype A Indian. The β-globin gene frameworks for the two alleles were 3 and 1, respectively. The Egyptian β-thalassemia allele represented an unusual haplotype and the β-globin gene was a framework 2 (Fig 1B). The Saudi Arabian allele was classified as a haplotype VII, framework 3a (3 Asian) gene (Fig 1C).

DNA sequence analysis. Complete DNA sequence analysis of all three exons, all intron/exon splice boundaries, and 5′ and 3′ flanking regions of the β-globin gene, revealed an identical C to A substitution within the IVS-2 acceptor consensus sequence at position −3 in both the Iranian and Egyptian patients (Fig 2). These mutations were detected by M13 sequencing and direct genomic sequencing, respectively, and were confirmed by oligonucleotide hybridization with normal and mutant probes (data not shown). In addition, the entire IVS-2 of the Iranian gene was sequenced, and no other abnormalities were observed. The other mutant allele in the Iranian patient is a common Indian frameshift mutation (+G) between codons 8 and 9.[14]

An analogous single base substitution at position −3, T to G, within the first IVS acceptor consensus sequence was identified in the Saudi Arabian patient by direct genomic sequencing. Direct sequence analysis of all three exons, intron 1, 5′ and 3′ flanking regions of the β-globin gene, and IVS-2 excluding nucleotides 100 to 630 revealed no other substitutions from a framework 3a gene. The other mutant allele, a haplotype II, framework 1 gene, had the common Mediterranean mutation at IVS-1 position 110.

In vitro expression studies. To examine the effects of the C to A mutation in the IVS-2 acceptor site consensus sequence on RNA processing, the β-globin gene was introduced into HeLa cells and RNA was isolated 48 hours after transfection for S1 nuclease mapping (Fig 3A).[4] Using a 443 nt *Eco* RI-*Rsa* I end-labeled probe, protected fragments of 65 nt and 320 nt were observed. The 65 nt fragment corresponds to RNA normally spliced at exon 3 whereas the 320 nt fragment reflects the use of a cryptic IVS-2 acceptor site at IVS-2 position 579.[4,15] This 320 nt fragment is only seen in transient expression assays using β-globin genes that contain a mutation 3′ to IVS-2 position 579 that affects

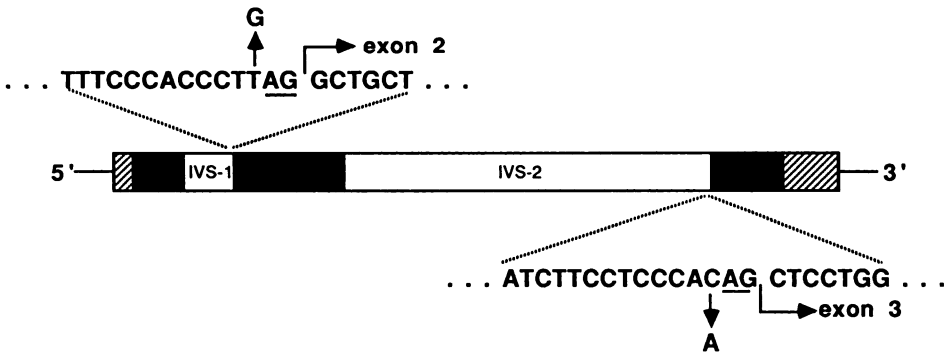


Fig 2. The β-globin IVS-2 acceptor splice site is diagrammed below. The β-globin gene and the IVS-1 acceptor splice site is shown above. Both consensus sequences include a 5′ pyrimidine stretch ($^{T}_{C}$)11 and the invariant AG dinucleotides (underlined). The new documented β-thalassemia mutations occur at positions −3 in both acceptor consensus sequences, a C to A transversion at IVS-2 and T to G transversion at IVS-1.

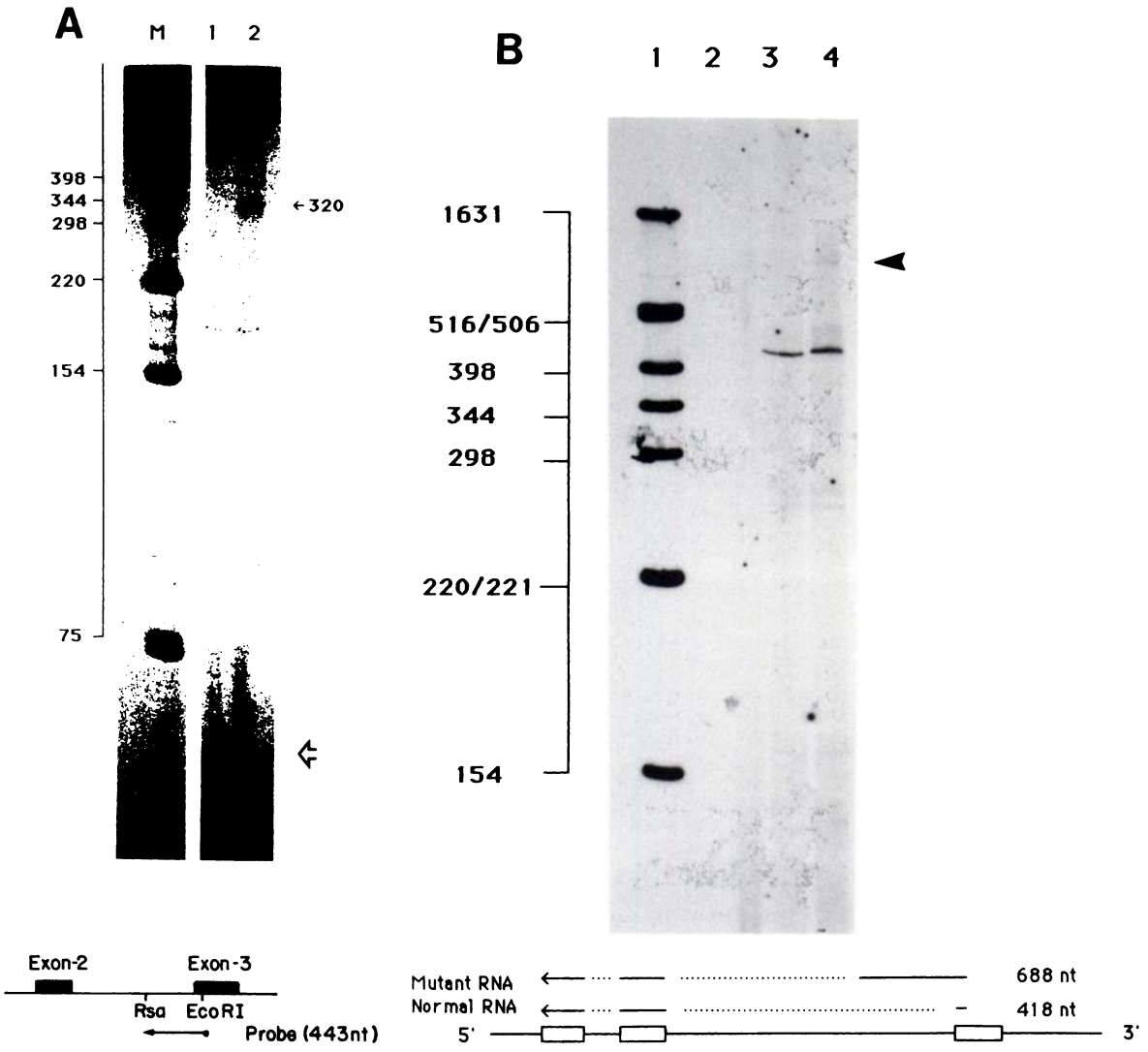


Fig 3. (A) Total cell RNA (25 γ) from HeLa cells transfected with the C to A mutant β-globin gene in the vector πSVplac was hybridized to the 5′ end-labeled probe shown at the bottom. Following S1 nuclease digestion the sample was electrophoresed in an 8% urea acrylamide gel. The normal protected fragment (65 nt) is indicated by the open arrow on the right. The novel protected fragment from the β-thalassemia gene is 320 nt. Lanes: M, marker; 1, normal β-globin gene; 2, mutant β-globin gene. (B) RNA processing of the same C to A mutant β-globin gene construct examined by primer extension analysis with an oligonucleotide primer located in exon 3 at the *Eco*RI site. Lanes: 1, *Hinf*I digested pBR322; 2, tRNA; 3, normal β-globin gene; 4, mutant β-globin gene. The normal 418 nt cDNA product is seen in both lanes 3 and 4 and diagrammed below the figure. Arrow at 688 nt indicates the abnormal extended product seen only in lane 4.

RNA splicing at the acceptor site of IVS-2. In these cases the processed RNA contains an extended exon 3. RNA splicing of IVS-1 appeared entirely normal (data not shown). The total level of β-globin RNA produced in the transient expression system was not markedly different from that of a normal β-globin gene.

The products of RNA processing of the IVS-2 mutant were also analyzed by primer extension studies (Fig 3B). RNA isolated from transfected HeLa cells was annealed to an oligonucleotide primer complementary to sequences in exon 3 from the *Eco*RI site and extending 5′. The cDNA product reverse transcribed from the normal gene is 418 nucleotides in length. The cDNA synthesized from the gene containing a mutant IVS-2 acceptor site is 688 nucleotides long, consistent in size with use of the cryptic splice site at position 579 in IVS-2. Although little abnormal extended product was observed, it was seen in three separate experiments and its quantity was similar to that expected from the S1 analysis.

Taken as a whole, these data demonstrate a low level use of a cryptic site in IVS-2 at position 579. Although the fragment size of the 320 nt fragment found by S1 analysis has an error of ±10 nt, no other potential cryptic acceptor site lies within 30 nucleotides on either side of position IVS-2, nt 579.

DISCUSSION

RNA splicing mutations in the β-globin gene represent close to one half of all documented β-thalassemia mutations.[16] Six of these RNA splicing defects are due to nucleo-

tide changes in the invariant dinucleotides, six are due to changes within IVS-1 donor consensus sequences, while eight are due to sequence changes within introns or exons that create a splice signal or enhance a cryptic splice site, which is used as an alternative or in addition to existing sites (see reference 16 for review). To date, however, no convincing evidence of mutations that alter the flanking consensus sequences in the IVS-1 acceptor site, the IVS-2 donor site, or IVS-2 acceptor site has been presented.

The development of the direct genomic sequencing technique has provided a rapid means for screening uncharacterized β-thalassemia alleles for new mutations. Examination of numerous genes from ethnic groups not previously studied for β-globin gene defects, for example, Iranian, Egyptian and Saudi Arabian (as in this study), is feasible. A new class of RNA splicing defect was identified based on this approach.

In both the Iranian and Egyptian patients, we identified an identical single base substitution adjacent to the invariant nucleotides AG of the IVS-2 acceptor splice site. The presence of the same mutation in different ethnic groups and on two different chromosomal backgrounds with regard to β-gene frameworks is best explained by independent origins of the mutation to the same critical nucleotide.[8] In both instances, normal mRNA is synthesized. However, a small fraction of initial RNA transcript is aberrantly processed. A previous mutation in an American black that altered the invariant AG pair to GG in the same IVS-2 acceptor site produced abnormal splicing through use of a cryptic acceptor within IVS-2 at position 579.[17,18] Normal splicing was completely abolished thus producing a severe β-thalassemia phenotype. Based on this splicing pattern, use of the same cryptic acceptor site at position 579 in IVS-2 was predicted and confirmed by S1 analysis and primer extension studies for the C to A change (Fig 3). Another mutation in the same IVS-2 acceptor site (AG to CG) was reported without functional studies.[19]

Tables computing the frequencies of nucleotides appearing at each position within the consensus sequence[1] have shown that C appears at position −3 of intron/exon boundaries 65% of the time. A T residue appears 31% of the time at this site, while A is observed 4% of the time and a G residue only 1% of the time. Based on the frequency at which a G residue appears at position −3 (1%) one would predict that the T to G substitution in the Saudi Arabian patient also decreases normal splicing. The preference of a pyrimidine at this −3 position is evident from these two mutations.

The S1 nuclease and primer extension data suggest that the IVS-2 consensus acceptor mutation is very mild. These data are reminiscent of those obtained in transient expression studies of the β^E gene, ie, minimal reduction in total β-globin synthesis and abnormally spliced RNA accounting for a small fraction of total spliced RNA.[20] Moreover, two independent origins of the IVS-2 positions −3 mutation and observation of the IVS-1 position −3 mutation strongly indicate that these mutations produce β-thalassemia through an effect, though mild, on RNA splicing.

The Iranian patient with the −3 IVS-2 defect in one β-globin gene and a frameshift mutation in the other β-globin gene surprisingly has a severe phenotype. Although one might surmise that the −3 IVS-2 defect is more severe in vivo than is the case in HeLa cells, a severe thalassemia has been observed by us in a similar situation. An Asian Indian child carrying the cap site (+1) mutation in one β-globin gene and a frameshift mutation in the other β-globin gene had β-thalassemia major[12] even though the cap site mutation is clinically silent in four family members who are carriers (mean MCV, 81.5 fL; mean HbA_2, 2.5%). A homozygote for the cap site mutation has an MCV of 71 fL and HbA_2 of 4.1%.

In a series of similar IVS-1 consensus mutants at positions 1, 5, and 6 in the IVS-1 splice site,[4] a gradient effect with respect to levels of correctly spliced RNA species is observed. At one extreme, the IVS-1 position 1 mutation completely inactivates the 5′ splice site resulting in no normally spliced mRNA. The position 5 mutant produces one half as much RNA as the normal β-globin gene, while the position 6 mutation produces almost normal levels of RNA.[4] This gradient effect is explainable by the efficiency with which the splicing machinery can discriminate between normal and cryptic splice sites. For example, three novel cryptic splice sites are used by the position 1 mutant. These same three sites are used by the position 5 and 6 mutants, although at lesser frequencies. The resulting variable level of normal mRNA is thus due to the reduced discriminatory ability of the processing enzymes. The presence of preferred nucleotides or stronger splicing machinery binding sites takes precedence over activation of cryptic sites. The difference in levels of abnormally spliced mRNA may also be due to the location of nucleotides within a secondary structure and subsequent recognition by processing enzymes. The AG pair hence appears at a more critical position in the ribonucleoprotein complex than the nucleotide at −3.

Thus, although the splice junction consensus sequences are well established, the mechanism of accurate splice-site selection remains unclear. Identification of additional consensus mutants through naturally occurring defects or via in vitro mutagenesis are needed to define sequence-dependent recognition or involvement in assembly of ribonucleoprotein complexes.

ACKNOWLEDGMENT

We thank Emily Pasterfield for expert assistance in manuscript preparation.

ADDENDUM

Gonzalez-Redondo et al have recently reported the occurrence of the C-A mutation at position −3 to the IVS-2/exon 3 splice junction in a black American with β-thalassemia.[21] Since the nt substitution was present in a haplotype I, framework 1 background, it represents a probable third independent origin of this mutation.

REFERENCES

1. Mount SM: A catalogue of splice junction sequences. Nucleic Acids Res 10:459, 1982

2. Shapiro MB, Senapathy P: RNA splice junctions of different classes of eukaryoties: Sequence statistics and functional implications in gene expression. Nucleic Acids Res 15:7155, 1987

3. Breathnach R, Benoist C, O'Hare K, Gannon F, Chambon P: Ovalbumin gene: Evidence for a leader sequence in mRNA and

DNA sequences at the exon-intron boundaries. Proc Natl Acad Sci USA 75:4853, 1978

4. Treisman R, Orkin SH, Maniatis T: Specific transcription and RNA splicing defects in five cloned β-thalassemia genes. Nature 302:591, 1983

5. Rees DJG, Rizza CR, Brownlee GG: Haemophilia B caused by a point mutation in a donor splice junction of the human factor IX gene. Nature 316:643, 1985

6. DiLella AG, Marvit J, Lidsky AS, Guttler F, Woo SLC: Tight linkage between a splicing mutation and a specific DNA haplotype in phenylketonuria. Nature 322:799, 1986

7. Beldjord C, Lapoumeroulie C, Pagnier J, Benabadji M, Bank A, Labie D: A novel β-thalassemia gene with a single base mutation in the conserved polypryimidine sequence at the 3' end of intron 2. Blood 70:71a, 1987 (suppl)

8. Wong C, Antonarakis SE, Goff SC, Orkin SH, Boehm CO, Kazazian HH Jr: On the origin and spread of β-thalassemia: Recurrent observation of four mutations in different ethnic groups. Proc Natl Acad Sci USA 83:6529, 1986

9. Orkin SH, Kazazian HH Jr, Antonarakis SE, Goff SC, Boehm CD, Sexton JP, Waber PG, Giardina PJV: Linkage of β-thalassemia mutations and β-globin gene polymorphisms with DNA polymorphisms in human β-globin gene cluster. Nature 296:627, 1982

10. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463, 1977

11. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA: Primer-directed enzymatic amplifications of DNA with a thermostable DNA polymerase. Science 239:487, 1988

12. Wong C, Dowling CE, Saiki RK, Higuchi R, Erlich HA, Kazazian HH Jr: Characterization of β-thalassemia mutations using direct genomic sequencing of amplified single copy DNA. Nature 330:384, 1987

13. Orkin SH, Kazazian HH Jr: The mutation and polymorphism of the human β-globin gene and its surrounding DNA. Ann Rev Genet 18:131, 1984

14. Kazazian HH, Jr, Orkin SH, Antonarakis SE, Sexton JP, Boehm CD, Goff SC, Waber PG: Molecular characterization of seven β-thalassemia mutations in Asian Indians. EMBO J 3:593, 1984

15. Dobkin C, Pergolizzi RG, Bahre P, Bank A: Abnormal splice in a mutant β-globin gene not at the site of the mutation. Proc Natl Acad Sci USA 80:1184, 1983

16. Kazazian HH Jr, Boehm CD: Molecular basis and prenatal diagnosis of β-thalassemia. Blood 72:1107, 1988

17. Antonarakis SE, Orkin SH, Cheng TC, Scott AF, Sexton JP, Trusko S, Charache S, Kazazian HH Jr: β-thalassemia in American blacks: Novel mutations in the TATA box and an acceptor splice site. Proc Natl Acad Sci USA 81:1154, 1984

18. Atweh GF, Anagnou NP, Forget BG, Kaufman RE: β-thalassemia resulting from a single nucleotide substitution in an acceptor splice site. Nucleic Acids Res 13:777, 1985

19. Padanilam BJ, Huisman THJ: The β^{o}-thalassemia in an American black family is due to a single nucleotide substitution in the acceptor splice junction of the second intervening sequence. Am J Hematol 22:259, 1986

20. Orkin SH, Kazazian HH Jr, Antonarakis SE, Ostrer H, Goff SC, Sexton JP: Abnormal RNA processing due to the exon mutation of the β^{E}-globin gene. Nature 300:768, 1982

21. Gonzalez-Redondo JM, Stoming TA, Lanclos KD, Gu YC, Kutlar A, Kutlar F, Nakatsuji T, Deng B, Han IS, McKie VC, Huisman THJ: Clinical and genetic heterogeneity in black patients with homozygous β-thalassemia from the southeastern United States. Blood 72:1007, 1988